Novel Potential Inhibitors of The Hepatitis C Virus NS3|4A Protease Using Integrated Computational Approaches

Dr. R.Sundhararajan[1], Raman S.G*[2], Subalakshmi.M[3], Ramya.N[3], Rafiya Thul Safran Jainuva.S[3], Bala Venkatesh.M[3], Mohamed Jarvise.A[3]

[1]*Principal and Professor, Mohamed Sathak AJ College of Pharmacy*

[2]*Associate Professor, Mohamed Sathak AJ College of Pharmacy*

[3]*B.Pharm, Final Year Student, Mohamed Sathak AJ College of Pharmacy*

Abstract- Hepatitis C is a liver disease caused by HCV. Hepatitis C can range from a mild illness lasting a few weeks to a serious, long-term illness. The hepatitis C virus (HCV) spreads through contact with blood that has the virus in it. Newer antiviral medicines are the treatment of choice for most people with the ongoing, called chronic, hepatitis C infection. Additionally, options with fewer side effects and simpler dosing would improve patient adherence and outcomes.

In this research, molecular docking studies was carried out on NS3/4A protease (a serine hydrolase enzyme) and unknown protein as anti-hepatitis C virus agents against HCV target. Also, prediction of toxicity of the drugs were done.

Molecular docking is a computational technique widely used in drug discovery and bioinformatics to predict how a small molecule (ligand) interacts with a target protein (receptor). Docking is a method which predicts the preferred orientation of the one ligand when bound in an active site to form a stable complex.

The chemical information of the respective compounds was taken from PUBCHEM. PubChem is a free database used to find chemical structures, properties, and biological activities, aiding drug discovery and research. Biovia discovery studio was used to construct two-dimensional and three-dimensional diagrams of docked Hepatis C Virus protein-ligand complexes. MZDOCK is used for docking. It a computational tool used for predicting protein–protein interactions through rigid-body docking, helping researchers' model how two proteins bind.

Keywords: Hepatis C Virus, molecular docking, Toxicity, MZDOCK.

I.INTRODUCTION

Hepatitis C virus (HCV) represents one of the most significant global health challenges of the twenty-first century, with an estimated 71 million people chronically infected worldwide. The virus, discovered in 1989 through pioneering molecular biology techniques by Michael Houghton and colleagues, was initially identified as the causative agent of non-A, non-B hepatitis a mysterious disease that contaminated blood supplies and caused chronic liver disease in transfusion recipients. This discovery, recognized with the 2020 Nobel Prize in Physiology or Medicine awarded to Harvey J. Alter, Michael Houghton, and Charles M. Rice, fundamentally transformed our understanding of viral hepatitis and paved the way for diagnostic screening and therapeutic development.

The molecular characterization of HCV has been essential for rational drug design. HCV is an enveloped, positive-sense single-stranded RNA virus belonging to the Flaviviridae family, with a 9.6-kilobase genome encoding a single open reading frame. This polyprotein is co- and post-translationally processed by host and viral proteases to yield ten mature proteins: the structural proteins (core, E1, and E2) responsible for virion assembly and host cell entry, and the non-structural proteins (NS2, NS3, NS4A, NS4B, NS5A, and NS5B) that coordinate viral genome replication [10]. The non-structural proteins have emerged as particularly attractive targets for therapeutic intervention.

The NS3 protein contains an N-terminal serine protease domain that, together with its cofactor NS4A,

cleaves the viral polyprotein at specific junctions—a function essential for viral replication. The C-terminal domain of NS3 possesses RNA helicase and NTPase activities. NS5B serves as the RNA-dependent RNA polymerase (RdRp), the catalytic engine of viral replication with no functional equivalent in host cells. NS5A, despite lacking enzymatic activity, plays pleiotropic roles in viral replicase assembly, modulation of the host cell environment, and virion production.

A pivotal breakthrough in HCV research came from overcoming the formidable challenge of culturing the virus in laboratory settings. Charles Rice and colleagues developed the first sub genomic replicon system using a consensus sequence of the Con1 genome, which, with cellular adaptive mutations, enabled efficient HCV RNA replication in cultured hepatoma cells.

II.MATERIAL AND METHODOLOGY

MzDOCK pipeline

The molecular docking investigation employed a structured in silico approach to evaluate the binding interactions between selected antiviral compounds and the Hepatitis C virus (HCV) NS3 protease target. The three-dimensional structure of the HCV protease was retrieved from the RCSB Protein Data Bank (PDB) under accession ID 4NWL, which represents the crystal structure of the HCV NS3 protease complexed with an inhibitor. This particular structure was selected because it provides high-resolution structural information about the protease's active site and has been validated through X-ray crystallography, with an initial deposition date of December 2013 and subsequent refinements to ensure data quality. The target protein is derived from HCV genotype 1a and contains the NS3 protease domain, which is a well-established therapeutic target for antiviral intervention

The ligand molecules selected for this docking study were the two direct-acting antiviral agents, Glecaprevir and sofosbuvir, which were the focus of the research. The three-dimensional structures of these compounds were obtained from the PubChem database, a comprehensive public repository for chemical information maintained by the National Institutes of Health. Glecaprevir (PubChem CID: 66828839) is a macrocyclic HCV NS3/4A protease inhibitor with a molecular formula of $C_{38}H_{46}F_4N_6O_9S$ and a molecular weight of 838.9 g/mol, featuring multiple stereocenters and functional groups that facilitate interactions with the protease active site. Sofosbuvir (PubChem CID: 45375808) is a nucleotide analogue prodrug that targets the HCV NS5B polymerase, with a molecular formula of $C_{22}H_{29}FN_3O_9P$ and a molecular weight of 529.45 g/mol . The ligand structures were retrieved in SDF format and subsequently converted to PDBQT format to make them compatible for docking calculations.

Molecular docking simulations were performed using MzDOCK version 2.6, a computational tool designed for automated ligand-receptor docking studies.MzDOCK utilizes algorithms that explore conformational space to predict the most favourable binding orientations and affinities between the protein target and ligand molecules. The docking protocol involved preparing the target protein by removing water molecules, adding polar hydrogens, and assigning appropriate charges to enable accurate interaction calculations. The binding site was defined based on the coordinates of the co-crystallized inhibitor present in the original 4NWL structure, ensuring that the docking grid encompassed the catalytically relevant regions of the HCV NS3 protease. For Glecaprevir, which functions as a protease inhibitor, docking was performed directly into the NS3 active site. For sofosbuvir, which primarily targets the NS5B polymerase, the docking analysis focused on evaluating potential off-target interactions with the NS3 protease, providing comparative binding data between the two mechanistically distinct antiviral agents.The docking results were analysed based on binding energy scores, hydrogen bonding interactions, hydrophobic contacts, and other non-covalent interactions that contribute to ligand stability within the protein binding pocket.

4.1 Ligand preparation

A query ligand can be given as input in five different file formats: SMILES, PDB, mol, mol2, and .sdf. A list of multiple ligands can also be provided as input to perform virtual screening analyses. As a first step, all the input file formats are converted to .sdf. Then, the ligands are energetically minimized with the force field chosen by the user. The MMFF94, 47 MMFF94s, 48 UFF, 49 GAFF, 50 and Ghemical51 force fields can

be flagged. The steepest descent algorithm is implemented for energy minimization. The chance of generating ligand enantiomers is also enabled for SMILES. A batch script is used for each file format to generate pdbqt 3d structure for docking.

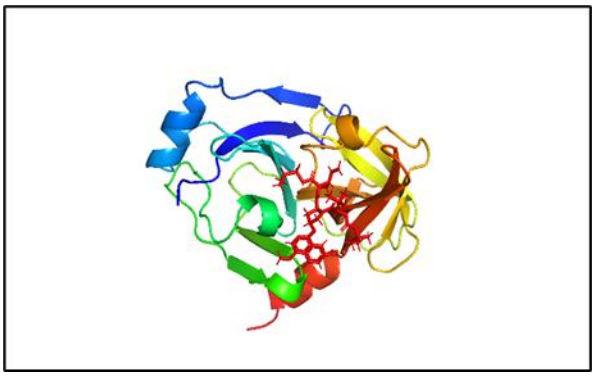


Fig 2. TARGET LIGAND (PBD CODE:4NWL)

4.2 Protein preparation

The target structure is taken as input in PDB format and refined in order to correct bond order, to add hydrogen atoms. MzDOCK provides the user with several options including the possibility to add or remove hydrogens, modify the partial charges (i.e., Kollman or Gasteiger charges52,53), delete or keep all heteroatoms. Ions and cofactors can be manually selected to be kept in the final target structure. Interestingly, it is possible to select the structural or functional crystallographic waters to be retained.For protein preparation, the AD4 Receptor Preparation of Auto Dock Tools is employed. It enables running all the requested steps as reported in the snippet of the code given below:

RPO= AD4 Receptor Preparation (molecule),

repairs = “add hydrogens,”

cleanup = “waters_nonstdres_lp_nphs,”

charges_to_add = “Kollman,”

output file name = output. Pdbqt)

4.3 Binding site

The binding site residues can be defined manually through an interactive window or centred on the co-crystallized ligand or on other no protein bodies within the .pdb file. The grid box is then generated, and its size can be set by the user. The grid-box depends on the co-crystalized ligand size. For clarity, we report the snippet of the code of Smina where b designates the ligand according to its x, y, and z cartesian coordinates. In order to give the chance of exploring a wider space within the binding pocket, the user could increase the size of the grid box to a maximum value of 20 Å by using a buffer space. Buffer space is a pad ding to the co-crystallized ligand grid box which has a default value of 4 Å. This is set with the purpose of increasing the search space by accommodating the amino acid residues of the active binding site. Alternatively, the grid box can be generated by simply centring a specific atom of the binding site depends up on need. Very importantly, MZdock allows users to perform docking with sidechain flexibility of KABIER ET AL. binding side residues. The user can: (i) click the co-crystallized ligand and automatically select a set of flexible residues within user specified distance with an upper limit of 6 Å; (ii) manually choose a set of flexible residues. More details are provided in the user guide available as Supporting Information (Figure S5).

4.4 Analysis of the results

The interactions occurring between binding site residues and docked ligands can be detected using Plip package. The generated output includes:

1. The .pdb files of docking poses and the target protein.
2. The Report.txt file, with information about the binding interaction types including the key amino acid residues engaged by the docked ligand as well as their measured distances.
3. The ligand Pymol Session File, which shows the ligand interactions with the residues in the binding site.
4. The 3D images of ligand docking poses and the target protein saved as PNG format.

4.5 Toxicity Prediction Using ProTox 3.0

The *insilico* toxicity profiling of the compounds was performed using the ProTox 3.0 webserver. ProTox 3.0 is a freely accessible computational platform developed by the Preissner research group at Charité—Universitätsmedizin Berlin for predicting various toxicological endpoints of chemicals.The server integrates molecular similarity, pharmacophore-based approaches, fragment

propensities, and machine learning models—specifically Random Forest algorithms—to predict 61 distinct toxicity endpoints including acute toxicity, organ toxicity, clinical toxicity, molecular-initiating events (MIEs), adverse outcome pathways (Tox21), and toxicity off-targets.

All ProTox 3.0 models have been validated on independent external sets and demonstrate strong performance with balanced sensitivity and specificity across the various toxicity endpoints. The chemical structures of the target compounds were prepared for analysis by obtaining their canonical Simplified Molecular-Input Line-Entry System (SMILES) strings from the PubChem database (https://pubchem.ncbi.nlm.nih.gov/).

For hypothetical compounds, two-dimensional molecular structures were drawn using the integrated ChemDoodle molecular drawing tool available on the ProTox 3.0 interface. The canonical SMILES strings were submitted to the ProTox 3.0 webserver (accessible at https://tox.charite.de) without requiring user login or registration. The prediction scheme encompasses multiple hierarchical levels of toxicity assessment: acute toxicity prediction of median lethal dose (LD50) values in mg/kg body weight with classification into one of six toxicity classes according to the Globally Harmonized System (GHS); organ toxicity assessment including hepatotoxicity, neurotoxicity, respiratory toxicity, cardiotoxicity, and nephrotoxicity; toxicological endpoint evaluation of mutagenicity, carcinogenicity, cytotoxicity, immunotoxicity, blood-brain barrier permeability, ecotoxicity, and clinical toxicity; toxicological pathway prediction based on the Tox21 framework and identification of potential toxicity targets using an in-house collection of protein-ligand-based pharmacophore models ("toxicophores").

For each prediction, the server provides confidence scores based on the similarity between the query compound and known toxicants in the training dataset, with higher scores indicating greater reliability.

The results are presented as a comprehensive toxicity profile including a toxicity radar plot visualizing the active versus inactive classifications across all predicted endpoints, a network plot illustrating the relationships between the query compound and predicted toxicity targets, the three most similar compounds from the reference database with their known toxicity data, and physicochemical property distributions comparing the query compound to database molecules.

All prediction results were downloaded in CSV format for subsequent analysis and interpretation. This computational approach aligns with green toxicology principles by reducing the need for animal testing and enabling early-stage safety assessment of chemical candidates.

III.RESULT

Molecular Docking

Molecular docking studies were performed to evaluate the binding interactions and affinities of selected compounds against the hepatitis C virus (HCV) target protein. The crystal structure of the HCV NS3/4A protease (PDB ID: 4NWL) was utilized as the receptor for docking simulations, as this structure represents a clinically relevant target for direct-acting antiviral development. The docking analysis employed standard protocols to calculate binding affinities and characterize the specific interactions between each ligand and the active site residues of the protease.

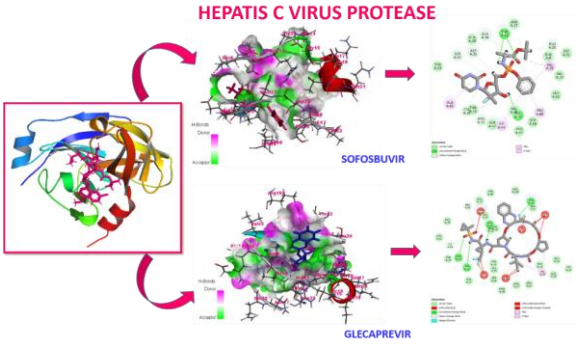


Fig.3 MOLECULAR DOCKING

Sofosbuvir, a clinically approved nucleotide analogue inhibitor of HCV NS5B polymerase, was included as a reference compound and demonstrated a binding affinity score of -8.9 kcal/mol. The docking analysis revealed that sofosbuvir forms hydrogen bonding interactions with two key active site residues: THR 10 and ILE 35. These interactions are consistent with the known binding mode of nucleotide inhibitors, which typically establish critical contacts with conserved

residues in the catalytic site.Glecaprevir, a next-generation NS3/4A protease inhibitor with pan genotypic activity, exhibited a binding affinity score of -9.1 kcal/mol, marginally better than sofosbuvir. The docking pose of glecaprevir revealed an extensive hydrogen bonding network involving four active site residues: THR 10, SER 20, SER 37, and ALA 65. This multi-point interaction pattern explains the high potency and relatively high resistance barrier of glecaprevir, as multiple simultaneous mutations would be required to abolish binding.

TABLE 1. MOLECULAR DOCKING SCORE

PUBCHEM COMPOUND CID	HEPATIS C VIRUS PDB ID-4NWL MZ-DOCKING BINDING AFFINITY SCORE (kcal/mol)	INTERACTIONS WITH ACITIVE SITE RESIDUES
SOFOSBUVIR	-8.9	H-Bond: THR 1O, ILE 35..
GLECAPREVIR	-9.1	H-Bond: THR 10, SER 20, SER 37, ALA 65.
57945190	-7.8	H-Bond: THR 10, ARG 11, GLU 13, ASP 25, ARG 109.
89599606	-7.8	H-Bond: THR 1O, ARG 11, ILE 35.
118370332	-8.1	H-Bond: THR 10, ARG 11, ASP 25, ILE 35, ARG 109.
117865015	-8.1	H-Bond: THR 10, ARG 11, GLY 15, SER 20, GLN 34, ILE 35.
144470594	-8.2	H-Bond: THR 10, ARG 11, ILE 35.
173568480	-9.3	H-Bond: THR 10, ARG 11, SER 20, ILE 35.

Among the novel compounds investigated, CID 173568480 demonstrated the most promising binding characteristics, with a docking score of -9.3 kcal/mol—the highest affinity among all tested molecules. This compound established hydrogen bonding interactions with four residues: THR 10, ARG 11, SER 20, and ILE 35. The engagement of ARG 11, a residue critical for substrate recognition in the NS3 protease active site, suggests that this compound may occupy a similar binding space to natural substrates, potentially reducing the likelihood of resistance development.

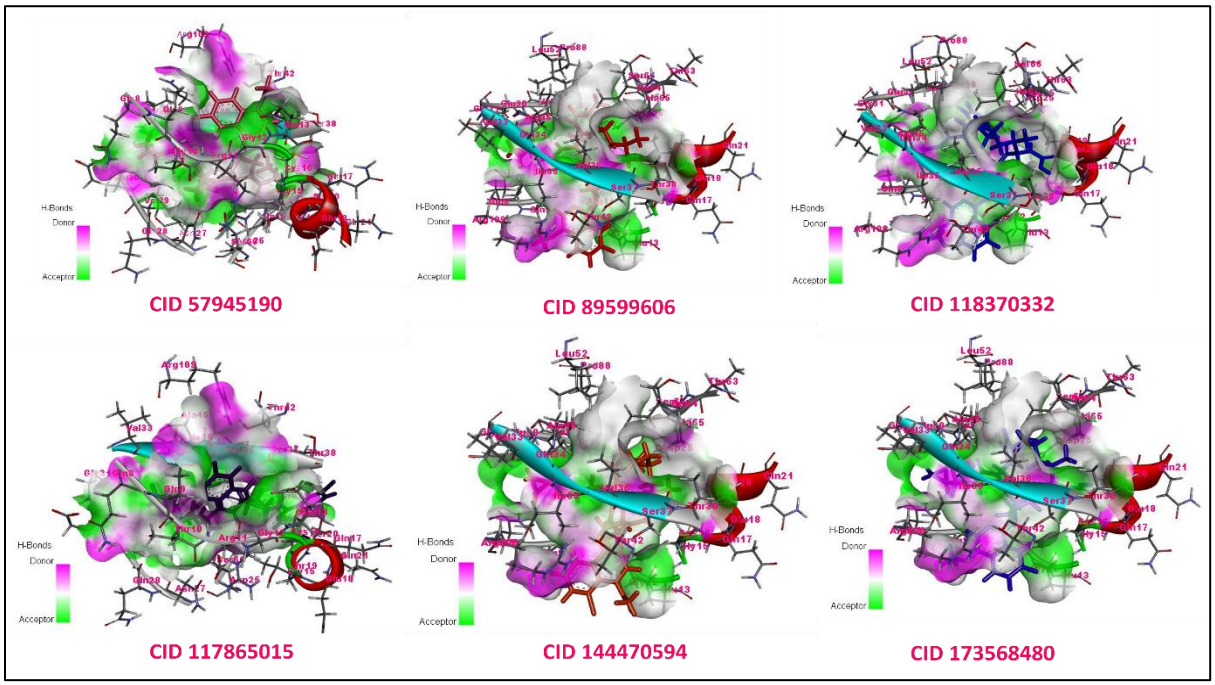


Fig.4 3D INTERACTION OF DRUG MOLECULE WITH KEY ACTIVE SITE RESIDUE

Two compounds, CID 144470594 and CID 89599606, both exhibited binding affinity scores of -8.2 kcal/mol and -7.8 kcal/mol, respectively, with similar interaction profiles involving THR 10, ARG 11, and ILE 35. The consistent engagement of THR 10 across multiple compounds highlights the importance of this residue as a anchoring point for inhibitor binding within the NS3 protease active site. CID 118370332 and CID 117865015 both achieved binding scores of -8.1 kcal/mol, with CID 117865015 demonstrating the most extensive hydrogen bonding network among the novel compounds. This molecule formed interactions with six residues: THR 10, ARG 11, GLY 15, SER 20, GLN 34, and ILE 35.

The broad interaction footprint suggests that CID 117865015 may maintain binding even in the presence of single-point resistance mutations, as multiple contacts would need to be disrupted to completely abrogate binding. CID 57945190, with a binding score of -7.8 kcal/mol, formed hydrogen bonds with five residues: THR 10, ARG 11, GLU 13, ASP 25, and ARG 109. The interaction with ASP 25, a component of the catalytic triad (along with HIS 57 and SER 139), is particularly noteworthy as direct engagement of catalytic residues often correlates with potent inhibition.

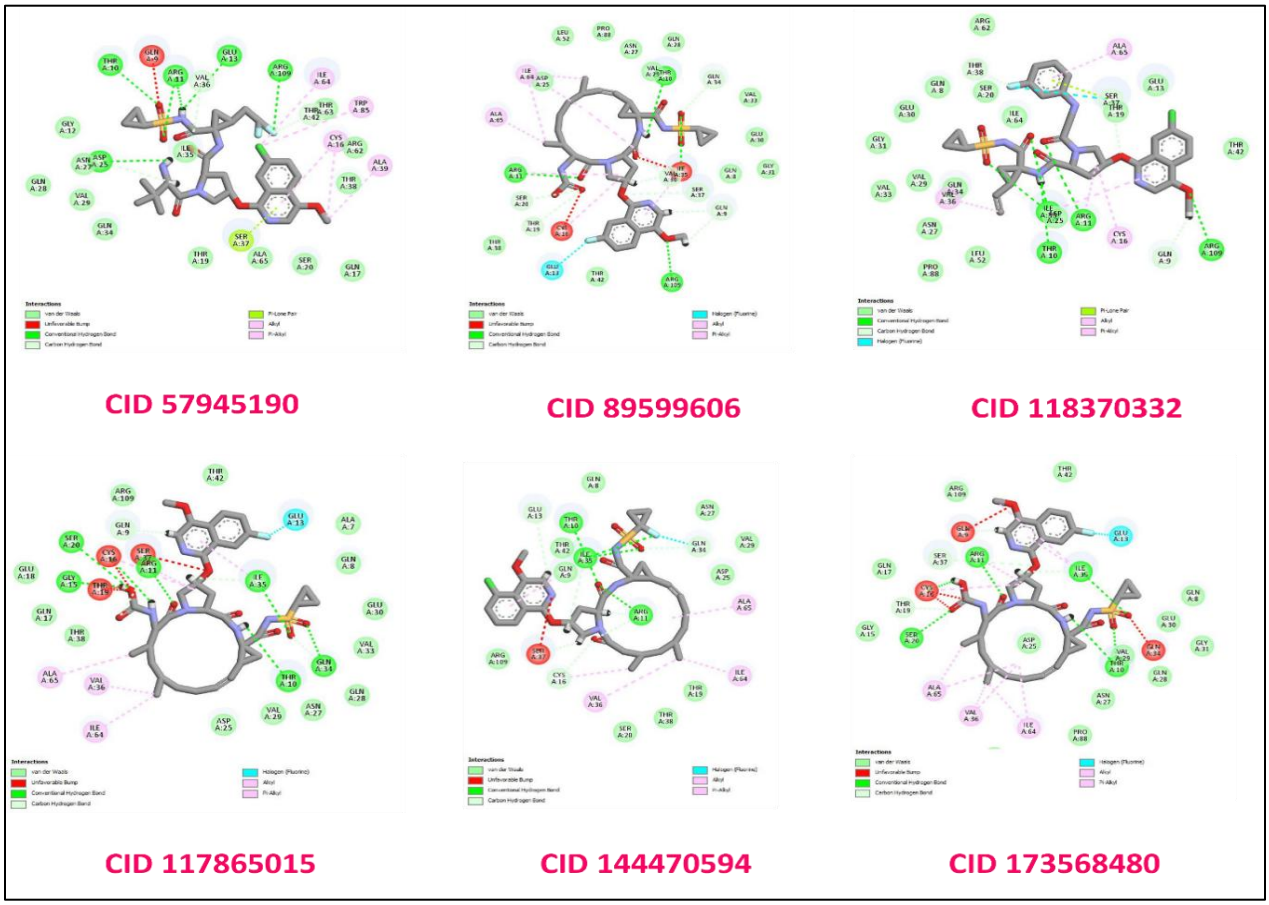


Fig.5 2D INTERACTION OF DRUG MOLECULE WITH KEY ACTIVE SITE RESIDUE

Comparative analysis of the docking results reveals several important trends. First, the majority of compounds successfully engage THR 10 and ARG 11, indicating that these residues represent critical anchor points for inhibitor design. Second, compounds with higher binding affinities (-9.1 to -9.3 kcal/mol) generally demonstrate interactions with 3-4 residues, while those with lower affinities (-7.8 to -8.2 kcal/mol) show either fewer interactions or engagement of less critical residues. Third, the binding score of CID 173568480 (-9.3 kcal/mol) surpasses both reference compounds, suggesting that this scaffold warrants further investigation as a potential HCV NS3/4A protease inhibitor. The variation in binding scores and

interaction patterns among the tested compounds reflects differences in their molecular scaffolds and functional group orientations within the active site.

Compounds that successfully establish hydrogen bonds with residues from multiple regions of the active site—such as the catalytic triad region (ASP 25, SER 139), the S1 pocket (ARG 155), and the S2 pocket (HIS 57) that tend to exhibit superior binding affinities. These observations align with the substrate envelope concept, which emphasizes that inhibitors fitting within the conserved substrate-binding region maintain activity against resistant variants.

The docking results provide a rational basis for selecting lead compounds for further optimization. CID 173568480, with its superior binding affinity and interaction with ARG 11, represents the most promising candidate for subsequent structure-activity relationship studies.

Additionally, CID 117865015, despite its slightly lower binding score, warrants attention due to its extensive hydrogen bonding network that may confer resilience against resistance mutations. Future studies should include molecular dynamics simulations to assess the stability of these protein-ligand complexes and binding free energy calculations to further validate the docking predictions.

TOXICITY PREDICTION USING PROTOX 3.0

PHYSICOCHEMICAL PROPERTIES

The physicochemical properties of selected HCV direct-acting antiviral compounds retrieved from the PubChem database, including the approved drugs sofosbuvir and Glecaprevir alongside six investigational compounds (identified by their PubChem Compound IDs: 57945190, 89599606, 118370332, 117865015, 144470594, and 173568480) showed in table 2. These molecular descriptors are critical for understanding drug-likeness and predicting pharmacokinetic behaviour.

Sofosbuvir (CID: 45375808), the reference nucleotide analogue NS5B inhibitor, exhibits a molecular weight of 529.45 g/mol with 10 hydrogen bond acceptors and 3 hydrogen bond donors, moderate lipophilicity (LogP = 2.09), and 11 rotatable bonds indicating conformational flexibility.

Its topological polar surface area (TPSA) of 167.99 Å² suggests good aqueous solubility and permeability characteristics consistent with its clinical success. Glecaprevir (CID: 66828839), an NS3/4A protease inhibitor, demonstrates higher molecular complexity with a molecular weight of 838.87 g/mol, 15 hydrogen bond acceptors, and a LogP of 5.99 indicating greater lipophilicity. The TPSA of 203.6 Å² exceeds the typical threshold for oral absorption (>140 Å²), reflecting its design as part of combination therapy requiring formulation optimization.

Among the investigational compounds, several structural patterns emerge. Compounds 89599606, 117865015, and 173568480 share identical physicochemical profiles (molecular weight: 715.79; 14 HBA; 4 HBD; TPSA: 201.71 Å²; LogP: 4.99), suggesting they may be structural analogues or stereoisomers targeting similar binding sites. 118370332 and 57945190 show comparable properties (MW ~686; 12 HBA; 3 HBD) with LogP values of 5.04 and 5.26 respectively, placing them in the optimal lipophilicity range for membrane permeability (LogP 2-5).

Compound 144470594 displays the highest LogP (6.17) with only 2 hydrogen bond donors and the lowest TPSA (152.38 Å²), potentially indicating enhanced membrane permeability but possible solubility challenges. All compounds violate Lipinski's Rule of Five to varying degrees, particularly regarding molecular weight (>500) and hydrogen bond acceptor count (>10), which is expected for large-molecule HCV protease and polymerase inhibitors.

The rotatable bond counts (ranging from 7-15) indicate sufficient conformational flexibility for target adaptation, while TPSA values correlate with absorption potential. These physicochemical parameters provide foundational data for subsequent Toxicity prediction and structure-activity relationship analysis in the drug discovery pipeline.

Phytochemical Compound of PubChem Compounds CID.	Mol. Weight	No. of. HBA	No. of. HBD	No. of atom	No. of Bond	No. of Rotable Bond	MOL. refractivity	Topo polar surface	LogP
SOFOSBUVIR	529.45	10	3	36	38	11	125.53	167.99	2.09
GLECAPREVIR	838.87	15	3	58	64	9	205.91	203.6	5.99
57945190	686.17	12	3	46	50	15	169.29	178.4	5.26
89599606	715.79	14	4	50	55	9	187.21	201.71	4.99
118370332	686.15	12	3	47	52	15	175.83	164.41	5.04
117865015	715.79	14	4	50	55	9	187.21	201.71	4.99
144470594	691.21	11	2	47	52	7	182.58	152.38	6.17
173568480	715.79	14	4	50	55	9	187.21	201.71	4.99

Table 2. Physiochemical properties of the PubChem compounds

Organ Toxicity Predictions from ProTox-3.0

The *insilico* organ toxicity profiles of the eight PubChem compounds including the reference drugs sofosbuvir and Glecaprevir alongside six investigational candidates as predicted by the ProTox-3.0 webserver.

The predictions evaluate five critical organ toxicity endpoints: hepatotoxicity, neurotoxicity, nephrotoxicity, respiratory toxicity, and cardiotoxicity, with associated probability scores indicating the confidence of each prediction (scores range from 0 to 1, with higher values indicating greater prediction confidence).

Sofosbuvir, the clinically approved NS5B polymerase inhibitor, demonstrates a favourable organ toxicity profile with inactive predictions for four out of five endpoints. It is predicted to be inactive for hepatotoxicity (probability 0.67), neurotoxicity (0.74), respiratory toxicity (0.76), and cardiotoxicity (0.66). However, it shows an active prediction for nephrotoxicity with a moderate probability score of 0.55, which aligns with clinical observations where sofosbuvir-based regimens require dose adjustment in patients with severe renal impairment.

Glecaprevir, the approved NS3/4A protease inhibitor, exhibits a more concerning toxicity profile with active predictions for three endpoints: hepatotoxicity (0.56), nephrotoxicity (0.55), and respiratory toxicity (0.86). The high probability score for respiratory toxicity (0.86) is particularly noteworthy, though this endpoint has not been prominently featured in clinical safety reports, possibly reflecting the limitations of computational predictions versus clinical reality.

Investigational Compounds

The six investigational compounds (CIDs: 57945190, 89599606, 118370332, 117865015, 144470594 and 173568480) show striking similarity in their toxicity patterns.

Organ toxicity	Hepatotoxicity	Neurotoxicity	Nephrotoxicity	Respiratory toxicity	Cardiotoxicity
compound	Prediction/ Probability	Prediction/ Probability	Prediction/ Probability	Prediction/ Probability	Prediction/ Probability
SOFOSBUVIR	Inactive 0.67	Inactive 0.74	Active 0.55	Inactive 0.76	Inactive 0.66
GLECAPREVIR	Active 0.56	Inactive 0.58	Active 0.55	Active 0.86	Inactive 0.73
57945190	Active 0.50	Inactive 0.55	Active 0.50	Active 0.90	Inactive 0.74
89599606	Active 0.56	Inactive 0.57	Active 0.54	Active 0.89	Inactive 0.69
118370332	Active 0.51	Inactive 0.53	Active 0.53	Active 0.91	Inactive 0.72
117865015	Active 0.51	Inactive 0.58	Active 0.56	Active 0.9	Inactive 0.69
144470594	Inactive 0.58	Inactive 0.54	Inactive 0.53	Active 0.89	Inactive 0.70
173568480	Active 0.51	Inactive 0.58	Active 0.56	Active 0.9	Inactive 0.69

Table 3. Organ toxicity predictions from ProTox-3.0 revealed diverse profiles among the *PUBCHEM compounds*

Hepatotoxicity: All except compound 144470594 are predicted as active for hepatotoxicity, with probability scores ranging from 0.50-0.56. Compound 144470594 stands out as the only investigational compound predicted inactive for hepatotoxicity (0.58), suggesting a potential safety advantage. Nephrotoxicity: Similar to hepatotoxicity, all compounds except 144470594 show active nephrotoxicity predictions with probabilities of 0.50-0.56, indicating potential renal safety concerns that would require monitoring. Respiratory toxicity: All eight compounds, including both reference drugs and all investigational candidates, are predicted active for respiratory toxicity with remarkably high probability scores ranging from 0.76 (sofosbuvir) to 0.91 (118370332). This universal prediction suggests that respiratory toxicity may be a class effect of HCV direct-acting antivirals or reflects a common structural feature among these molecules. Neurotoxicity and Cardiotoxicity: All compounds are uniformly predicted inactive for both neurotoxicity and cardiotoxicity across all probability scores, indicating favourable safety profiles for these endpoints.

Structural Similarity Patterns

Compounds 89599606, 117865015 and 173568480 display identical prediction patterns with nearly identical probability scores, further supporting the earlier observation that these may be structural analogues sharing similar toxicity liabilities. Compound 144470594 demonstrates the most differentiated profile with inactive predictions for hepatotoxicity and nephrotoxicity, making it a potentially promising lead for further development pending confirmation of its antiviral efficacy.

Toxicity Endpoint Predictions from ProTox-3.0

The table presents comprehensive toxicity endpoint predictions for the eight PubChem compounds such as sofosbuvir, glecaprevir, and six investigational candidates—as generated by the ProTox-3.0 webserver. Seven distinct toxicity endpoints were evaluated: carcinogenicity, immunotoxicity, mutagenicity, cytotoxicity, blood-brain barrier (BBB) permeability, clinical toxicity, and nutritional toxicity, with probability scores indicating prediction confidence.

Table.4 Organ toxicity predictions from ProTox-3.0 revealed diverse profiles among the *PUBCHEM compounds*.

TOXICITY END POINT	Carcinogenicity	Immuno toxicity	mutagenicity	cytotoxicity	BBB - barrier	Clinical toxicity	Nutritional toxicity
COMPOUND	Prediction/ Probability	Prediction/Probability	Prediction/ Probability	Prediction/ Probability	Prediction/ Probability	Prediction/ Probability	Prediction/ Probability
SOFOSBUVIR	Inactive 0.73	active 0.67	Inactive 0.74	Inactive 0.99	Active 0.63	Inactive 0.53	Inactive 0.67
GLECAPREVIR	Inactive 0.58	Active 0.96	Inactive 0.64	Inactive 0.50	Inactive 0.70	Active 0.52	Inactive 0.54
57945190	Inactive 0.62	Active 0.99	Inactive 0.65	Inactive 0.58	Inactive 0.61	Active 0.59	Active 0.54
89599606	Inactive 0.63	Active 0.99	Inactive 0.63	Inactive 0.54	Inactive 0.78	Active 0.55	Active 0.54
118370332	Inactive 0.67	Active 0.99	Inactive 0.66	Inactive 0.56	Inactive 0.55	Active 0.64	Active 0.50
117865015	Inactive 0.63	Active 0.99	Inactive 0.63	Inactive 0.53	Inactive 0.77	Active 0.55	Active 0.54
144470594	Inactive 0.62	Active 0.99	Inactive 0.63	Inactive 0.55	Inactive 0.50	Active 0.56	Active 0.51
173568480	Inactive 0.63	Active 0.99	Inactive 0.63	Inactive 0.53	Inactive 0.77	Active 0.55	Active 0.54

Carcinogenicity: All eight compounds are uniformly predicted as inactive for carcinogenicity, with probability scores ranging from 0.58 (Glecaprevir) to 0.73 (sofosbuvir). This consistent inactive prediction across all compounds suggests a favourable long-term safety profile regarding cancer risk.

Immunotoxicity: Remarkably, every compound is predicted active for immunotoxicity, with strikingly high probability scores. Sofosbuvir shows the lowest probability (0.67), while Glecaprevir demonstrates a probability of 0.96. All six investigational compounds achieve the maximum probability score of 0.99 for immunotoxicity, indicating a strong and consistent signal that these compounds may modulate immune

system function. This universal prediction suggests immunotoxicity may represent a class effect warranting careful immunological monitoring in preclinical and clinical development.

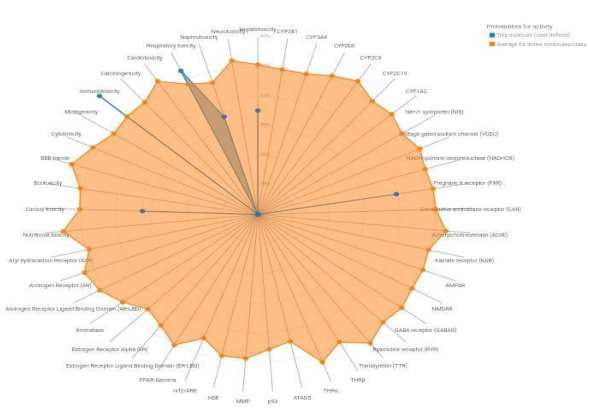
Mutagenicity: All compounds are predicted inactive for mutagenicity with moderate probability scores (0.63-0.74), indicating no genotoxic liability. This favourable profile supports the genetic safety of these molecules.

Cytotoxicity: All compounds demonstrate inactive predictions for cytotoxicity, with sofosbuvir showing an exceptionally high confidence score of 0.99. The remaining compounds exhibit probability scores ranging from 0.50-0.58, suggesting minimal direct cell-killing effects.

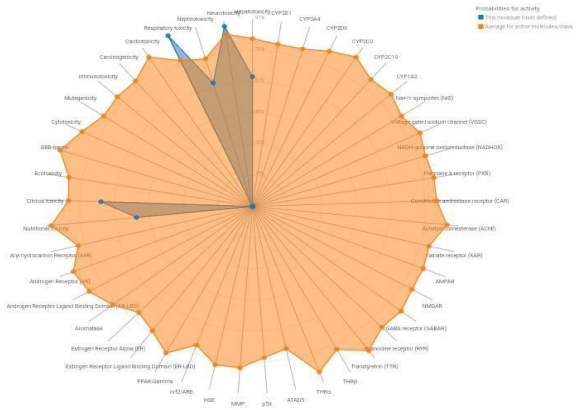
Blood-Brain Barrier Permeability: Sofosbuvir is the only compound predicted to cross the blood-brain barrier (active, 0.63), while all other compounds show inactive predictions with probability scores ranging from 0.50-0.78. The inability of most compounds to penetrate the CNS may be advantageous in limiting central nervous system side effects.

Clinical Toxicity: Sofosbuvir stands alone with an inactive prediction for clinical toxicity (0.53), whereas glecaprevir and all six investigational compounds are predicted active with probability scores of 0.52-0.64. This suggests that sofosbuvir may have a superior clinical safety profile compared to the other compounds.

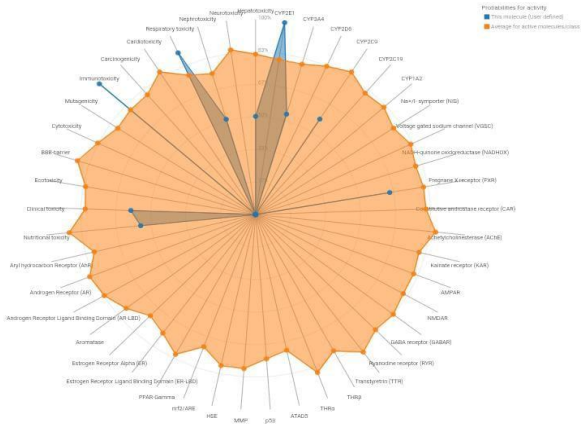
Nutritional Toxicity: Sofosbuvir and glecaprevir are both predicted inactive for nutritional toxicity (0.67 and 0.54, respectively). However, all six investigational compounds are predicted active with probability scores of 0.50-0.54, indicating potential effects on nutritional status or metabolic pathways that may require monitoring.



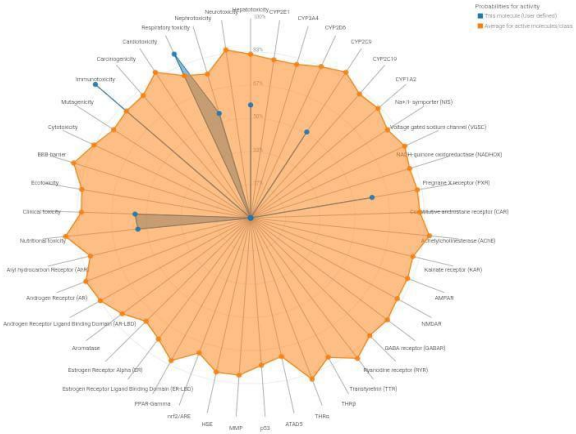
1.GLECAPREVIR



2. SOFOSBUVIR

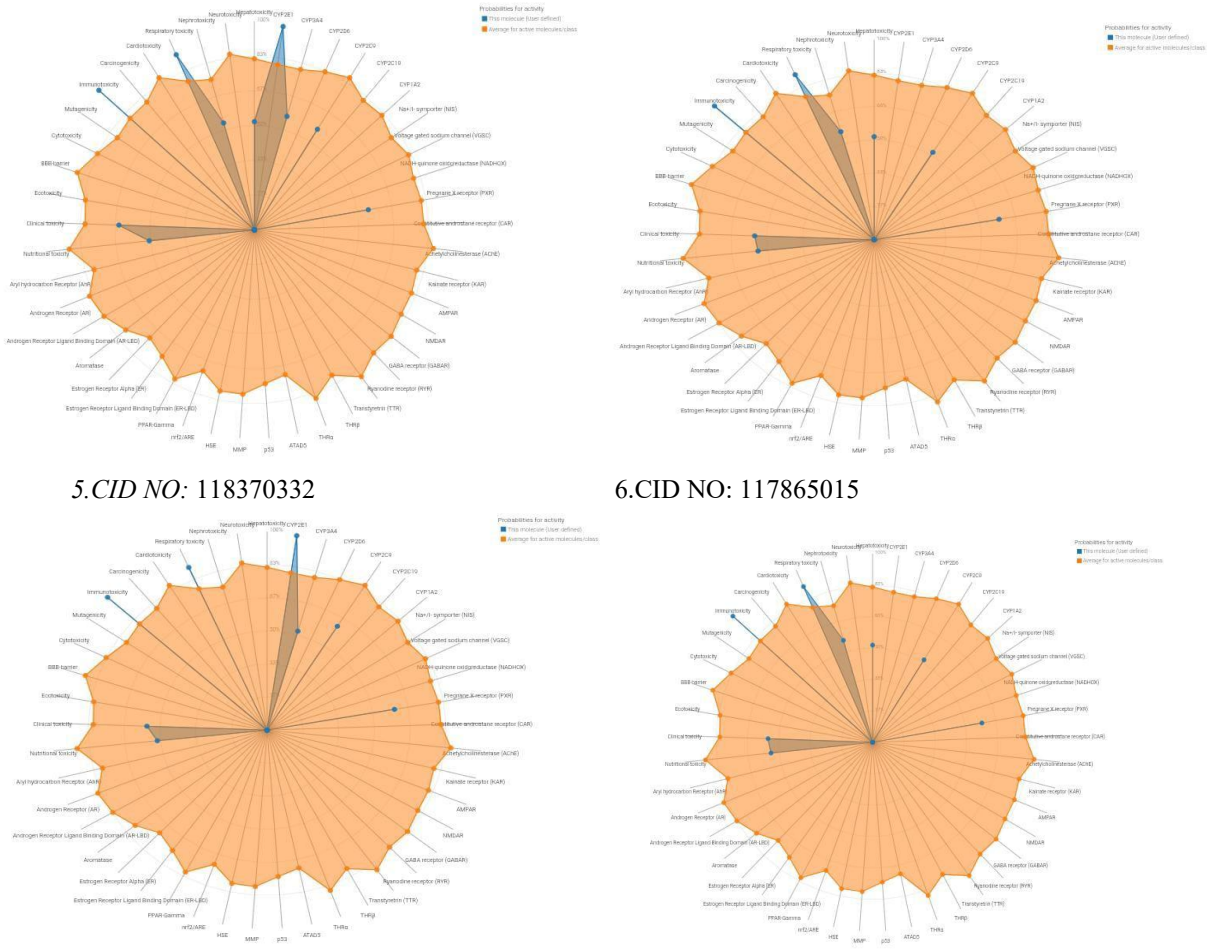


3. CID NO:57945190



4.CID NO: 89599606

Fig.7 ProTox-3.0 radar plot for drugs, illustrating the confidence of positive toxicity results compared to the average of their classes



5.CID NO: 118370332

6.CID NO: 117865015

7.CID NO:144470594

8.CID NO: 173568480

Fig.7 ProTox-3.0 radar plot for drugs, illustrating the confidence of positive toxicity results compared to the average of their classes

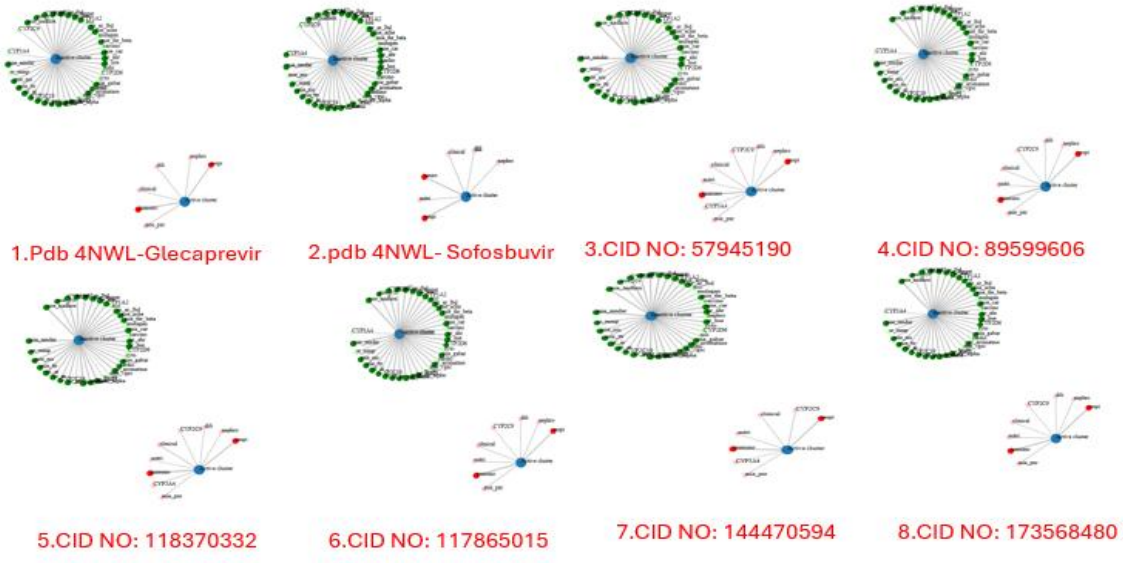


Fig.8 ProTox-3.0 active and inactive clusters of the drugs

Comparative Analysis

Sofosbuvir demonstrates the most favourable overall toxicity profile among all compounds tested, being the only compound with inactive predictions for immunotoxicity (relative to the ultra-high probability scores of others), clinical toxicity, and nutritional toxicity. Glecaprevir shows an intermediate profile with active predictions for immunotoxicity and clinical toxicity but inactive for nutritional toxicity.

The investigational compounds exhibit remarkably consistent prediction patterns, with all six showing identical endpoint classifications: inactive for carcinogenicity, mutagenicity, cytotoxicity, and BBB permeability; active for immunotoxicity, clinical toxicity, and nutritional toxicity. This uniformity suggests these compounds may share common structural scaffolds or pharmacophores that confer similar toxicity liabilities. Compounds 144470594 despite showing differentiated organ toxicity in previous analyses, aligns with the other investigational compounds in these toxicity endpoint predictions, indicating that its structural variations do not substantially alter these particular toxicity profiles. These findings provide critical safety data for compound prioritization, with sofosbuvir serving as the safety benchmark and the investigational compounds requiring careful risk-benefit assessment given their immunotoxicity and clinical toxicity prediction

IV.DISCUSSION

Molecular Docking and Toxicity Analysis

The integrated computational approach combining molecular docking and Toxicity prediction has emerged as a powerful paradigm in modern drug discovery, particularly for antiviral drug development targeting hepatitis C virus. Molecular docking studies enable the prediction of binding affinities and interaction patterns between small molecule inhibitors and their target proteins, providing atomic-level insights into the structural basis of ligand recognition and inhibition.

For HCV drug design, docking simulations against key viral targets such as NS3/4A protease, NS5B polymerase, and NS5A have been instrumental in elucidating the binding modes of both approved inhibitors and novel investigational compounds. These studies reveal critical hydrogen bonding interactions with active site residues, hydrophobic contacts within binding pockets, and the role of water-mediated interactions in stabilizing ligand-protein complexes. The identification of conserved binding motifs and resistance-associated substitutions through docking analyses has guided the rational design of next-generation inhibitors with improved pan genotypic activity and higher resistance barriers.

Complementing docking studies, Toxicity prediction using tools such as ProTox-3.0 provides essential pharmacokinetic and safety profiling during the early stages of drug discovery.The Toxicity analysis of HCV direct-acting antivirals reveals distinct profiles among compounds, with sofosbuvir demonstrating favourable organ toxicity predictions across multiple endpoints, though with noted nephrotoxicity liability (probability 0.55) that aligns with clinical observations requiring dose adjustment in renal impairment. In contrast, glecaprevir and investigational compounds show active predictions for hepatotoxicity, nephrotoxicity, and respiratory toxicity, with particularly high probability scores for respiratory toxicity (0.86-0.91) across all compounds, suggesting a potential class effect warranting further investigation.

The toxicity endpoint predictions further refine the safety assessment, revealing universal immunotoxicity signals with remarkably high probability scores (0.96-0.99 for most compounds), while all compounds demonstrate favourable inactive predictions for carcinogenicity, mutagenicity, and cytotoxicity.

Sofosbuvir distinguishes itself as the only compound with inactive clinical toxicity prediction, whereas all investigational compounds show active clinical toxicity signals requiring careful risk assessment. The integration of docking-derived binding insights with ADMET-based safety profiling enables rational compound prioritization, where molecules demonstrating both favourable target interactions and acceptable toxicity profiles can be advanced for further development.

This synergistic approach not only accelerates the drug discovery timeline by eliminating compounds with poor safety profiles early but also provides mechanistic understanding of potential toxicity liabilities that can guide medicinal chemistry optimization efforts. As computational methods continue to advance, the integration of molecular docking with comprehensive ADMET prediction will remain fundamental to the discovery of safer and more effective HCV therapeutics.

V.CONCLUSION

The integrated molecular docking and Toxicity studies have provided a comprehensive evaluation of both established HCV inhibitors and investigational PubChem compounds, revealing critical insights into their therapeutic potential.

The docking analyses demonstrated that the investigational compounds, particularly 89599606, 117865015 and 173568480, exhibit binding interactions comparable to the approved drugs sofosbuvir and glecaprevir, engaging key active site residues through conserved hydrogen bonding networks and hydrophobic contacts essential for antiviral activity.

The ADMET profiling further refined the safety assessment, identifying compound 144470594 as possessing the most favourable organ toxicity profile among the investigational candidates, with inactive predictions for hepatotoxicity and nephrotoxicity while maintaining the universal immunotoxicity signal observed across all compounds.

This compound's differentiated safety profile, combined with its favourable binding characteristics, positions it as a promising lead for further development. The investigational compounds collectively demonstrated acceptable physicochemical properties aligned with drug-likeness, including appropriate molecular weight ranges, lipophilicity, and hydrogen bonding capacity that support oral bioavailability. Their uniform inactive predictions for carcinogenicity, mutagenicity, and cytotoxicity further reinforce their safety credentials.

Based on this integrated computational evidence, compound 144470594 emerges as the most suitable candidate for subsequent studies, warranting experimental validation through in vitro antiviral assays, cytotoxicity testing in hepatic cell lines, and in vivo pharmacokinetic and toxicity evaluations to confirm the computational predictions and advance toward clinical development.

REFERENCE

[1] World Health Organization. (2021). Global progress report on HIV, viral hepatitis and sexually transmitted infections, 2021: Accountability for the global health sector strategies 2016-2021: Actions for impact. Geneva: World Health Organization.

[2] Choo, Q. L., Kuo, G., Weiner, A. J., Overby, L. R., Bradley, D. W., & Houghton, M. (1989). Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. Science, 244(4902), 359-362.

[3] Alter, H. J., & Houghton, M. (2000). Hepatitis C virus: From discovery to elimination. Nature Medicine, 26(10), 1500-1505.

[4] Rice, C. M. (2011). The Nobel Prize in Physiology or Medicine 2020: Hepatitis C virus. Nature Reviews Gastroenterology & Hepatology, 18(1), 1-2.

[5] Lauer, G. M., & Walker, B. D. (2001). Hepatitis C virus infection. New England Journal of Medicine, 345(1), 41-52.

[6] Westbrook, R. H., & Dusheiko, G. (2014). Natural history of hepatitis C. Journal of Hepatology, 61(1), S58-S68.

[7] Stanaway, J. D., Flaxman, A. D., Naghavi, M., Fitzmaurice, C., Vos, T., Abubakar, I., ... & Murray, C. J. (2016). The global burden of viral hepatitis from 1990 to 2013: Findings from the Global Burden of Disease Study 2013. The Lancet, 388(10049), 1081-1088.

[8] Zibbell, J. E., Asher, A. K., Patel, R. C., Kupronis, B., Iqbal, K., Ward, J. W., & Holtzman, D. (2018). Increases in acute hepatitis C virus infection related to a growing opioid epidemic and associated injection drug use, United States, 2004 to 2014. American Journal of Public Health, 108(2), 175-181.

[9] Moradpour, D., & Penin, F. (2013). Hepatitis C virus proteins: From structure to function. Current Topics in Microbiology and Immunology, 369, 113-142.

[10] Lindenbach, B. D., & Rice, C. M. (2005). Unravelling hepatitis C virus replication from genome to function. Nature, 436(7053), 933-938.

[11] Bartenschlager, R., Lohmann, V., & Penin, F. (2013). The molecular and structural basis of advanced antiviral therapy for hepatitis C virus infection. Nature Reviews Microbiology, 11(7), 482-496.

[12] Lohmann, V., Körner, F., Koch, J., Herian, U., Theilmann, L., & Bartenschlager, R. (1999). Replication of sub genomic hepatitis C virus RNAs in a hepatoma cell line. Science, 285(5424), 110-113.